WHAT IS CLAIMED IS:

1. A word importance calculation method for calculating the importance of words contained in a document set, whereby the difference between the word distribution in a subset of whole documents which consists of every document containing a specified word and the word distribution in the set of whole documents is used to calculate the importance of the word.

2. A word importance calculation method, as claimed in Claim 1, wherein:

said difference is determined by comparing the distance d between said subset and said set of whole documents with the distance d', or the estimated value of d', between another subset of documents which contain substantially the same number of words as said subset of documents and are randomly selected from said set of whole documents, and said set of whole documents.

3. A word importance calculation method, as claimed in Claim 2, wherein:

the distance d between the two document sets is calculated by using the word distribution in each document set, that is to say using the probability of occurrence of each word in each of said document set.

4. A word importance calculation method, as claimed in Claim 2, wherein:

if the number of documents containing said word is larger than a prescribed number, a preset number of documents are extracted from the said subset of whole documents by random sampling, and the difference between the extracted set of documents and said set of whole documents is used instead of the difference between the original subset of documents and the set of whole documents.

5. A document retrieval interface having a function to display on a screen words characterizing a document set, wherein the importance of each word occurring in the set of whole documents is calculated using the difference between the word distribution in the subset of whole documents containing the word and the word distribution in the set of whole documents, and the importance is brought to bear on the selection, arrangement or coloring of the words displayed on the screen.

6. A document retrieval interface having a function to display on a screen words characterizing a document set, wherein the importance of each word occurring in the document set obtained as a result of retrieval is calculated using the difference between the word distribution in the subset of documents out of the document set obtained as a result of that retrieval containing that word and the word distribution in the document set obtained as a result of that retrieval, and the importance is brought to bear on the

selection, arrangement or coloring of the words displayed on the screen.

7. A word dictionary construction method by extracting important words from a document set in accordance with rules given in advance, wherein the importance of each word occurring in a set of whole documents is calculated using the difference between a subset of whole documents containing the word and the word distribution in the set of whole documents, and words to be extracted are selected on the basis of that importance.

8. A word importance calculation method whereby;

a characteristic quantity of a document set containing a certain word and

a characteristic quantity of a randomly extracted document set of the size of said set are compared; and

the importance of said word is thereby calculated.

9. A word importance calculation method, as claimed in Claim 8, wherein:

the difference between the word distribution in said document set and the word distribution in the set of all documents is used as said characteristic quantity.